PCT/GB 2003 / 0 0 1 5 4 3

INVESTOR IN PEOPLE

PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

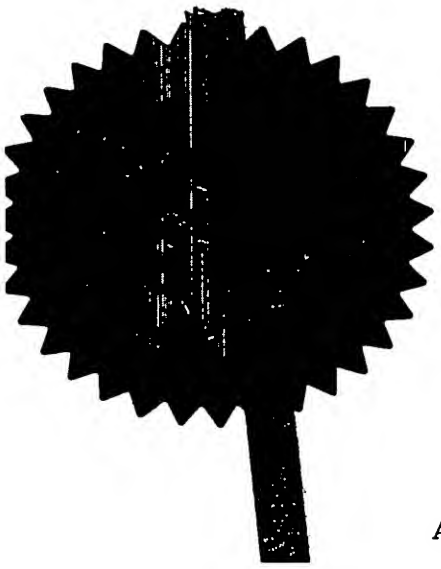
RECD 20 JUN 2003
WIPO PCT

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.



Signed R. Mahoney

Dated 16 APR 2003

Patents Form 1/77

Patents Act 1977
(Rule 16)

The Patent Office

1/77

05APR02 E708918-4 D00239
P01/7700 0.00-0207901.0

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

THE PATENT OFFICE
B
- 5 APR 2002
NEWPORT

The Patent Office
Cardiff Road
Newport
Gwent NP9 1RH

1. Your reference PC/JM/P12116GB

2. Patent application number *(The Patent Office will fill in this part)* 05 APR 2002 0207901.0

3. Full name, address and postcode of the or of each applicant *(underline all surnames)*

THE UNIVERSITY COURT OF THE UNIVERSITY
OF EDINBURGH
OLD COLLEGE
SOUTH BRIDGE
EDINBURGH EH8 9YL
UNITED KINGDOM

04001590001
Patents ADP number *(if you know it)*

If the applicant is a corporate body, give the country/state of its incorporation

4. Title of the invention "SCHIZOPHRENIA ASSOCIATED GENE (IV)"

5. Name of your agent *(if you have one)* CRUIKSHANK & FAIRWEATHER

"Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*

19 ROYAL EXCHANGE SQUARE
GLASGOW G1 3AE
UNITED KINGDOM

Patents ADP number *(if you know it)* 547002

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and *(if you know it)* the or each application number

Country	Priority application number *(if you know it)*	Date of filing *(day / month / year)*

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application	Date of filing *(day / month / year)*

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer 'Yes' if:*

a) any applicant named in part 3 is not an inventor, or
b) there is an inventor who is not named as an applicant, or
c) any named applicant is a corporate body.
See note (d))

Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form	
Description	29
Claim(s)	–
Abstract	–
Drawing(s)	7 + 7

10. If you are also filing any of the following, state how many against each item.

Priority documents	
Translations of priority documents	
Statement of inventorship and right to grant of a patent *(Patents Form 7/77)*	
Request for preliminary examination and search *(Patents Form 9/77)*	
Request for substantive examination *(Patents Form 10/77)*	
Any other documents *(please specify)*	

11. I/We request the grant of a patent on the basis of this application.

Signature Cruikshank & Fairweather Date 4/4/02

CRUIKSHANK & FAIRWEATHER 4 APRIL 2002

12. Name and daytime telephone number of person to contact in the United Kingdom DR. P. CHAPMAN 0141-221-5767

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

a) If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505.

b) Write your answers in capital letters using black ink or you may type them.

c) If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.

d) If you have answered 'Yes' Patents Form 7/77 will need to be filed.

e) Once you have filled in the form you must remember to sign and date it.

f) For details of the fee and ways to pay please contact the Patent Office.

SCHIZOPHRENIA ASSOCIATED GENE (IV)

The present invention relates to the identification of a gene which has been disrupted in a patient diagnosed as suffering from schizophrenia, as well as proteins encoded by the gene and antibodies thereto and to uses of such products as medicaments for treating schizophrenia and/or affective psychosis. The invention also relates to methods for diagnosing patients suffering or predisposed to schizophrenia and/or affective psychosis, as well as screens for developing novel treatment regimes for schizophrenia and/or affective psychosis.

Schizophrenia and affective psychoses such as Bipolar Affective Disorder are common and debilitating psychiatric disorders. Despite a wealth of information on the epidemiology, neuroanatomy and pharmacology of the illness, it is uncertain what molecular pathways are involved and how impairments in these affect brain development and neuronal function. Despite an estimated heritability of 60-80%, very little is known about the number or identity of genes involved in these psychoses. Although there has been recent progress in linkage and association studies, especially from genome-wide scans, these studies have yet to progress from the identification of susceptibility loci or candidate genes to the full characterisation of disease-causing genes (Berrettini, 2000).

The cloning of breakpoints in patients with chromosome abnormalities (translocations, inversions etc.) has proved instrumental in the identification of many disease genes

(e.g. Duchenne Muscular Dystrophy, Retinoblastoma, Wilm's Tumour, Familial Polyposis Coli, Fragile-X Syndrome, Polycystic Kidney Disease, many leukaemias and, very recently, a candidate speech and language disorder gene (Lai et al, 2001)). Such studies assume that the chromosomal breakpoints give rise to the clinical symptoms by either directly disrupting gene sequences or perturbing gene expression. In the latter case, it is considered that the breakpoint can still affect expression of a gene at a distance of at least up to 1Mb (Kleinjan & van Heyningen, 1998). In the same way that gene-trap mutagenesis can be used to identify disrupted mouse genes (Brennan & Skarnes, 1999), the physical "flag" created by a cytogenetic breakpoint provides a geographical pointer for the disease locus.

It is amongst the objects of the present invention to provide a gene and/or protein postulated to be involved with the development and/or symptoms associated with schizophrenia and/or affective psychosis.

The present invention is based on the characterisation of a balanced reciprocal translocation between chromosomes 9 and 14, t(9;14)(q34;q13) in a mother with schizophrenia and her daughter with schizophrenia co-morbid with mild learning disability. A brain transcription factor gene, *NPAS3*, is shown to be disrupted by the translocation at 14q13. Without wishing to be bound by theory, the present inventors hypothesis is that the disruption of this gene is responsible for the psychotic symptoms exhibited by the

mother and daughter.

In a first aspect the present invention provides use of a polynucleotide fragment comprising the *NPAS3* gene or a fragment, derivative or homologue thereof for the manufacture of a medicament for treating schizophrenia and/or affective psychosis in a subject.

In another aspect the present invention provides use of a polypeptide fragment encoded by the *NPAS3* gene, or a fragment, derivative or homologue thereof for the manufacture of a medicament for treating schizophrenia and/or affective psychosis in a subject.

Schizophrenia and/or affective psychosis as used herein relates to schizophrenia, as well as other affective psychoses such as those listed in "The ICD-10 Classification of Mental and Behavioural Disorders" World Health Organization, Geneva 1992. Categories F20 to F29 inclusive includes Schizophrenia, schizotypal and delusional disorders. Categories F30 to F39 inclusive are Mood (affective) disorders that include bipolar affective disorder and depressive disorder. Mental Retardation is coded F70 to F79 inclusive. The Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DMS-IV). American Psychiatric Association, Washington DC. 1994. Include al conditions coded 295.xx (Schizophrenia and Other Psychotic Disorders) and 296.xx (Major Depressive Disorders and Bipolar Disorders). Mental Retardation is coded 315, 317, 318 and 319.

The human form of *NPAS3* has previously been identified and is found in the public database under accession numbers AB054575 and AF164438, with the differences due to alternative splicing and all forms are encompassed within the present invention.

Thus, references herein to the *NPAS3* gene are understood to relate to the sequences identified in Figures 1 and 3 and references to the NPAS3 protein sequence are understood to relate to the sequences identified in Figures 2 and 4.

In certain jurisdictions claims to methods of treatment are permissible and so the skilled reader will appreciate that the *NPAS3* gene, or fragment, derivative or homologue thereof; or NPAS3 protein, or functionally active fragment, derivative, or homologue thereof, may be administered to an individual as a method of treating an individual with schizophrenia and/or affective psychosis.

"Polynucleotide fragment" as used herein refers to a chain of nucleotides such as deoxyribose nucleic acid (DNA) and transcription products thereof, such as RNA. Naturally, the skilled addressee will appreciate the whole naturally occurring human genome is not included in the definition of polynucleotide fragment.

The polynucleotide fragment can be isolated in the sense that it is substantially free of biological material with which the whole genome is normally associated *in vivo*. The isolated polynucleotide fragment may be cloned to provide a recombinant molecule comprising the

polynucleotide fragment. Thus, "polynucleotide fragment includes double and single stranded DNA, RNA and polynucleotide sequences derived therefrom, for example, subsequences of said fragment and which are of any desirable length. Where a nucleic acid is single stranded then both a given strand and a sequence or reverse complementary thereto is within the scope of the present invention.

In general, the term "expression product" or "gene product" refers to both transcription and translation products of said polynucleotide fragments. When the expression or gene product is a "polypeptide" (i.e. a chain or sequence of amino acids displaying a biological activity substantially similar (eg. 98%, 95%, 90%, 80%, 75% activity) to the biological activity of the protein), it does not refer to a specific length of the product as such. Thus, the skilled addressee will appreciate that "polypeptide" encompasses *inter alia* peptides, polypeptides and proteins. The polypeptide if required, can be modified *in vivo* and *in vitro*, for example by glycosylation, amidation, carboxylation, phosphorylation and/or post-translational cleavage.

The present invention further provides a recombinant or synthetic polypeptide for the manufacture of reagents for use as therapeutic agents in the treatment of schizophrenia and/or affective psychosis. In particular, the invention provides pharmaceutical compositions comprising the recombinant or synthetic polypeptide

together with a pharmaceutically acceptable carrier therefor.

The present invention further provides an isolated polynucleotide fragment capable of specifically hybridising to a related polynucleotide sequence from another species. In this manner, the present invention provides probes and/or primers for use in *ex vivo* and/or *in situ* detection and expression studies. Typical detection studies include polymerase chain reaction (PCR) studies, hybridisation studies, or sequencing studies. In principle any specific polynucleotide sequence fragment from the identified sequences may be used in detection and/or expression studies. The skilled addressee understands that a specific fragment is a fragment of the sequence which is of sufficient length, generally greater than 10, 12, 14, 16 or 20 nucleotides in length, to bind specifically to the sequence, under conditions of high stringency, as defined herein, and not bind to unrelated sequences, that is sequences from elsewhere in the genome of the organism other than an allelic form of the sequence or non-homologous sequences from other organisms.

"Capable of specifically hybridising" is taken to mean that said polynucleotide fragment preferably hybridises to a related or similar polynucleotide sequence in preference to unrelated or dissimilar polynucleotide sequences.

The invention includes polynucleotide sequence(s) which are capable of specifically hybridising to an polynucleotide fragment as described herein or to a part

thereof without necessarily being completely complementary or reverse complementary to said related polynucleotide sequence or fragment thereof. For example, there may be at least 50%, or at least 75%, at least 90%, or at least 95% complementarity. Of course, in some cases the sequences may be exactly reverse complementary (100% reverse complementary) or nearly so (e.g. there may be less than 10, typically less than 5 mismatches). Thus, the present invention also provides anti-sense or complementary nucleotide sequence(s) which is/are capable of specifically hybridising to the disclosed polynucleotide sequence. If a specific polynucleotide is to be used as a primer in PCR and/or sequencing studies, the polynucleotide must be capable of hybridising to related nucleic acid and capable of initiating chain extension from 3' end of the polynucleotide, but not able to correctly initiate chain extension from unrelated sequences.

If a polynucleotide sequence of the present invention is to be used in hybridisation studies to obtain or identify a related sequence from another organism the polynucleotide sequence should preferably remain hybridised to a sample polynucleotide under stringent conditions. If desired, either the test or sample polynucleotide may be immobilised. Generally the test polynucleotide sequence is at least 10, 14, 20 or at least 50 bases in length. It may be labelled by suitable techniques known in the art. Preferably the test polynucleotide sequence is at least 200 bases in length and may even be several kilobases in

length. Thus, either a denatured sample or test sequence can be first bound to a support. Hybridization can be effected at a temperature of between 50 and 70°C in double strength SSC (2xNaCl 17.5g/l and sodium citrate (SC) at 8.8g/l) buffered saline containing 0.1% sodium dodecyl sulphate (SDS). This can be followed by rinsing of the support at the same temperature but with a buffer having a reduced SSC concentration. Depending upon the degree of stringency required, and thus the degree of similarity of the sequences, such reduced concentration buffers are typically single strength SSC containing 0.1%SDS, half strength SSC containing 0.1%SDS and one tenth strength SSC containing 0.1%SDS. Sequences having the greatest degree of similarity are those the hybridisation of which is least affected by washing in buffers of reduced concentration. It is most preferred that the sample and inventive sequences are so similar that the hybridisation between them is substantially unaffected by washing or incubation in standard sodium citrate (0.1 x SSC) buffer containing 0.1%SDS.

Oligonucleotides may be designed to specifically hybridise to NPAS3 nucleic acid. They may be synthesised, by known techniques and used as primers in PCR or sequencing reactions or as probes in hybridisations designed to detect the presence of mutated or normal NPAS3 in a sample. The oligonucleotides may be labelled by suitable labels known in the art, such as, radioactive labels, chemiluminescent labels or fluorescent labels and

the like.

The term "oligonucleotide" is not meant to indicate any particular length of sequence and encompasses nucleotides of preferably at least 10b (e.g. 10b to 1kb) in length, more preferably 12b-500b in length and most preferably 15b to 100b.

The oligonucleotides may be designed with respect to any of the sequences described herein and may be manufactured according to known techniques. They may have substantial sequence identity (e.g. at least 50%, at least 75%, at least 90% or at least 95% sequence identity) with one of the strands shown therein or an RNA equivalent, or with a part of such a strand. Preferably such a part is at least 10, at least 30, at least 50 or at least 200 bases long. It may be an open reading frame (ORF) or a part thereof.

Oligonucleotides which are generally greater than 30 bases in length should preferably remain hybridised to a sample polynucleotide under one or more of the stringent conditions mentioned above. Oligonucleotides which are generally less than 30 bases in length should also preferably remain hybridised to a sample polynucleotide but under different conditions of high stringency. Typically the melting temperature of an oligonucleotide less than 30 bases may be calculated according to the formula of; 2°C for every A or T, plus 4°C for every G or C, minus 5°C. Hybridization may take place at or around the calculated melting temperature for any particular oligonucleotide, in

6 x SSC and 1% SDS. Non specifically hybridised oligonucleotides may then be removed by stringent washing, for example in 3 x SSC and 0.1% SDS at the same temperature. Only substantially similar matched sequences remain hybridised i.e. said oligonucleotide and corresponding test nucleic acid.

When oligonucleotides of generally less than 30 bases in length are used in sequencing and/or PCR studies, the melting temperature may be calculated in the same manner as described above. The oligonucleotide may then be allowed to anneal or hybridise at a temperature around the oligonucleotides calculated melting temperature. In the case of PCR studies the annealing temperature should be around the lower of the calculated melting temperatures for the two priming oligonucleotides. It is to be appreciated that the conditions and melting temperature calculations are provided by way of example only and are not intended to be limiting. It is possible through the experience of the experimenter to vary the conditions of hybridisation and thus anneal/hybridise oligonucleotides at temperatures above their calculated melting temperature. Indeed this can be desirable in preventing so-called non-specific hybridisation from occurring.

It is possible when conducting PCR studies to predict an expected size or sizes of PCR product(s) obtainable using an appropriate combination of two or more oligonucleotides, based on where they would hybridise to the sequences described herein. If, on conducting such a

PCR on a sample of DNA, a fragment of the predicted size is obtained, then this is predictive that the DNA encodes a homologous sequence from a test organism.

Proteins for all the applications described herein can be produced by cloning the gene for example into plasmid vectors that allow high expression in a system of choice e.g. insect cell culture, yeast, animal cells, bacteria such as *Escherichia coli*. To enable effective purification of the protein, a vector may be used that incorporates an epitope tag (or other "sticky" extension such as His6) onto the protein on synthesis. A number of such vectors and purification systems are commercially available.

The polynucleotide fragment can be molecularly cloned into a prokaryotic or eukaryotic expression vector using standard techniques and administered to a host. The expression vector is taken up by cells and the polynucleotide fragment of interest expressed, producing protein.

It will be understood that for the particular polypeptides embraced herein, natural variations such as may occur due to polymorphisms, can exist between individuals or between members of the family. These variations may be demonstrated by (an) amino acid difference(s) in the overall sequence or by deletions, substitutions, insertions, inversions or additions of (an) amino acid(s) in said sequence. All such derivatives showing the recognised activity are included within the scope of the invention. For example, for the purpose of

the present invention conservative replacements may be made between amino acids within the following groups:

(I) Alanine, serine, threonine;

(II) Glutamic acid and aspartic acid;

(III) Arginine and leucine;

(IV) Asparagine and glutamine;

(V) Isoleucine, leucine and valine;

(VI) Phenylalanine, tyrosine and tryptophan

Moreover, recombinant DNA technology may be used to prepare nucleic acid sequences encoding the various derivatives outlined above.

As is well known in the art, the degeneracy of the genetic code permits substitution of bases in a codon resulting in a different codon which is still capable of coding for the same amino acid, e.g. the codon for amino acid glutamic acid is both GAT and GAA. Consequently, it is clear that for the expression of polypeptides from nucleotide sequences described herein or fragments thereof, use can be made of a derivative nucleic acid sequence with such an alternative codon composition different from the nucleic acid sequences shown in the Figures.

The polynucleotide fragments of the present invention are preferably linked to regulatory control sequences. Such control sequences may comprise promoters, operators, inducers, enhancers, silencers, ribosome binding sites, terminators etc. Suitable control sequences for a given host may be selected by those of ordinary skill in the art.

A polynucleotide fragment according to the present invention can be ligated to various expression controlling sequences, resulting in a so called recombinant nucleic acid molecule. Thus, the present invention also includes an expression vector containing an expressible nucleic acid molecule. The recombinant nucleic acid molecule can then be used for the transformation of a suitable host.

Specific vectors which can be used to clone nucleic acid sequences according to the invention are known in the art (e.g. Rodriguez, R.L. and Denhadt, D.T., Edit., Vectors: a survey of molecular cloning vectors and their uses, Butterworths, 1988, or Jones *et al.*,Vectors: Cloning Applications: Essential Techniques (Essential techniques series), John Wiley & Son. 1998).

The methods to be used for the construction of a recombinant nucleic acid molecule according to the invention are known to those of ordinary skill in the art and are inter alia set forth in Sambrook, et al. (Molecular Cloning: a laboratory manual Cold Spring Harbour Laboratory, 1989).

The present invention also relates to a transformed cell containing the polynucleotide fragment in an expressible form. "Transformation", as used herein, refers to the introduction of a heterologous polynucleotide fragment into a host cell. The method used may be any known in the art, for example, direct uptake, transfection transduction or electroporation (Current Protocols in Molecular Biology, 1995. John Wiley and Sons Inc.). The

heterologous polynucleotide fragment may be maintained through autonomous replication or alternatively, may be integrated into the host genome. The recombinant nucleic acid molecules preferably are provided with appropriate control sequences compatible with the designated host which can regulate the expression of the inserted polynucleotide fragment, e.g. tetracycline responsive promoter, thymidine kinase promoter, SV-40 promoter and the like.

Suitable hosts for the expression of recombinant nucleic acid molecules may be prokaryotic or eukaryotic in origin. Hosts suitable for the expression of recombinant nucleic acid molecules may be selected from bacteria, yeast, insect cells and mammalian cells.

In another aspect the present invention also relates to a method of diagnosing schizophrenia and/or affective psychosis or susceptibility to schizophrenia and/or affecive psychosis in an individual, wherein the method comprises determining if the *NPAS3* gene in the individual has been disrupted by a mutation or chromosomal rearrangement.

The methods which may be employed to elucidate such a mutation or chromosomal rearrangement are well known to those of skill in the art and could be detected for example using PCR or in hybridisation studies using suitable probes which could be designed to span an identified mutation site or in the *NPAS3* gene, such as the breakpoint identified by the present inventors and described herein.

Once a particular polymorphism or mutation has been identified it may be possible to determine a particular course of treatment. For example it is known that some forms of treatment work for some patients, but not all. This may in fact be due to mutations in the *NPAS3* gene or surrounding sequence, and it may therefore be possible to determine a treatment strategy using current therapies, based on a patient's genotype.

It will be appreciated that mutations in the gene sequence or controlling elements of a gene, eg. a promoter and/or enhance can have subtle effects such as affecting mRNA splicing/stability/activity and/or control of gene expression levels, which can also be determined.

Also the relative levels of RNA can be determined using for example hybridisation or quantitative PCR as a means to determine if the *NPAS3* gene has been disrupted.

Moreover the presence and/or levels of the *NPAS3* gene products themselves can be assayed by immunological techniques such as radioimmunoassay, Western blotting and ELISA using specific antibodies raised against the gene products. The present invention also therefore relates to antibodies specific for a *NPAS3* gene product and uses thereof in diagnosis and/or therapy.

A further aspect of the present invention therefore provides antibodies specific to the polypeptides of the present invention or epitopes thereof. Production and purification of antibodies specific to an antigen is a matter of ordinary skill, and the methods to be used are

clear to those skilled in the art. The term antibodies can include, but is not limited to polyclonal antibodies, monoclonal antibodies (mAbs), humanised or chimeric antibodies, single chain antibodies, Fab fragments, $F(ab')_2$ fragments, fragments produced by a Fab expression library, anti-idiotypic (anti-Id) antibodies, and epitope binding fragments of any of the above. Such antibodies may be used in modulating the expression or activity of the particular polypeptide, or in detecting said polypeptide *in vivo* or *in vitro*.

Using the sequences disclosed herein, it is possible to identify related sequences in other animals, such as mammals, with the intention of providing an animal model for psychiatric disorders associated with the improper functioning of the nucleotide sequences and proteins of the present invention. Once identified, the homologous sequences can be manipulated in several ways common to the skilled person in order to alter the functionality of the nucleotide sequences and proteins homologous to those of the present invention. For example, "knock-out" animals may be created, that is, the expression of the genes comprising the nucleotide sequences homologous to those of the present invention may be reduced or substantially eliminated in order to determine the effects of reducing or substantially eliminating the expression of such genes. Alternatively, animals may be created where the expression of the nucleotide sequences and proteins homologous to those of the present invention are upregulated, that is,

the expression of the genes comprising the nucleotide sequences homologous to those of the present invention may be increased in order to determine the effects of increasing the expression of these genes. In addition to these manipulations, substitutions, deletions and additions may be made to the nucleotide sequences encoding the proteins homologous to those of the present invention in order to effect changes in the activity of the proteins to help elucidate the function of domains, amino acids, etc. in the proteins. Furthermore, the sequences of the present invention may also be used to transform animals to the manner described above. The manipulations described above may also be used to create an animal model of schizophrenia and/or affective psychosis associated with the improper functioning of the nucleotide sequences and/or proteins of the present invention in order to evaluate potential agents which may be effective for combatting psychotic disorders, such as schizophrenia and/or affective psychosis.

Thus, the present invention also provides for screens for identifying agents suitable for preventing and/or treating schizophrenia and/or affective psychosis associated with disruption or alteration in the expression of the *NPAS3* gene and/or its gene products. Such screens may easily be adapted to be used for the high throughput screening of libraries of compounds such as synthetic, natural or combinatorial compound libraries.

Thus, the *NPAS3* gene products according to the present invention can be used for the *in vivo* or *in vitro* identification of novel ligands or analogs thereof. For this purpose binding studies can be performed with cells transformed with nucleotide fragments according to the invention or an expression vector comprising a polynucleotide fragment according to the invention, said cells expressing the *NPAS3* gene products according to the invention.

Alternatively also the *NPAS3* gene products according to the invention as well as ligand-binding domains thereof can be used in an assay for the identification of functional ligands or analogs for the *NPAS3* gene products.

Methods to determine binding to expressed gene products as well as *in vitro* and *in vivo* assays to determine biological activity of gene products are well known. In general, expressed gene product is contacted with the compound to be tested and binding, stimulation or inhibition of a functional response is measured.

Thus, the present invention provides for a method for identifying ligands for *NPAS3* gene products, said method comprising the steps of:

a) introducing into a suitable host cell a polynucleotide fragment according to the invention;

b) culturing cells under conditions to allow expression of the polynucleotide fragment;

c) optionally isolating the expression product;

d) bringing the expression product (or the host cell from step b)) into contact with potential ligands which will possibly bind to the protein encoded by said polynucleotide fragment from step a);

e) establishing whether a ligand has bound to the expressed protein; and

f) optionally isolating and identifying the ligand.

As a preferred way of detecting the binding of the ligand to the expressed protein, also signal transduction capacity may be measured.

Compounds which activate or inhibit the function of *NPAS3* gene products may be employed in therapeutic treatments to activate or inhibit the polypeptides of the present invention.

The present invention will now be further described by way of Example and with reference to the Figures which show:

Table 1 probes used in FISH

Figure 1 shows the nucleic acid sequence of *NPAS3* spliceform 1;

Figure 2 shows the protein sequence of NPAS3 spliceform;

Figure 3 shows the nucleic acid sequence of *NPAS3* spliceform 2;

Figure 4 shows the protein sequence of NPAS3 spliceform;

Figure 5 shows an ideogram representation of the balanced translocation in the patient relating to this invention.

Figure 6 shows the genomic arrangement of the NPAS3 gene including the position of the observed breakpoint; and

Figure 7 shows potential functional consequences of the disruption to NPAS3 gene : dominant-negative activity.

Lymphocyte extraction and metaphase chromosome preparation

Lymphocytes were extracted from 7mls of patient blood (for storage and generation of EBV-transformed cell lines) using density gradient separation (Histopaque-1077, Sigma). In order to generate metaphase-arrested chromosomes for cytogenetic analysis, 0.8mls of patient blood were cultured for 71hrs in medium containing phytohaemagglutinin (Peripheral Blood Medium, Sigma). The short-term cultures were treated with colcemid for one hour followed by a conventional fixing procedure. Fixed chromosomes were dropped onto microscope slides and stored for 1 week prior to use in FISH experiments.

Selection of YAC clones for FISH probe synthesis

YAC clones were selected from the Whitehead/MIT maps of the cytogenetic interval within which the 14q12 breakpoint was adjudged to lie; accessed through the Genome Database (http://gdbwww.gdb.org/). YACs were obtained from the HGMP Resource Centre, Hinxton, Cambridge, UK. DNA was prepared from these clones by standard methods and

amplified by PCR using primers designed against the consensus sequence elements within the archetypal *Alu* repeat (Breen et al, 1992). This "Alu-PCR" gives a representative spread of non-repetitive sequence over the full length of the YAC and yields a better FISH probe synthesis than native YAC DNA. Alu-PCR was performed using the expand long template PCR kit (Roche).

Fluorescence in situ hybridisation (FISH) protocol

Pooled Alu-PCR/purified BAC or cosmid DNA was labelled with digoxigenin or biotin by nick translation and hybridised to metaphase spreads using standard FISH methods. Slides were counterstained with DAPI in Vectashield antifade solution (Vector laboratories). A Zeiss Axioskop fluorescence microscope with a chroma number 81000 multi spectral filter set was used to examine slides. Images were captured using Vysis smart capture extension running within IP Lab spectrum. The probe signal hybridised to the chromosomal region of interest indicating whether the probe was centromeric or telomeric to the breakpoint. This result dictated the choice of further DNA probes for FISH. The process continued until a probe signal was split into three distinct signals (on normal 14, der14 and der 9), indicating that it spanned the breakpoint.

Translation of YAC FISH results onto the BAC map

Markers present within the YAC were identified from the Whitehead Institute map. These were placed within a BAC contig on the UCSC GoldenPath Draft Genome Browser (http://genome.ucsc.edu/goldenPath/hgTracks.html) database and BAC clone order was verified by consultation of the Washington University FPC database (http://www.genome.wustl.edu/gsc/human/Mapping/index.shtml). RPCI-11 library BAC clones were ordered from BACPAC Resources, Children's Hospital, Oakland, Cal., USA and subsequently taken through FISH analysis as above.

Fine FISH mapping of the breakpoint with cosmid clones

PCR products corresponding to regions in or near hnPAS3 exons 4, 5 and 6 were obtained using the following primers under standard PCR conditions (Exon 4-*i* ACAACCATTCTGGGAACAGC, Exon 4-*ii* GTGTAGGGAAAGCCATCCAA, Exon 5-*i* TCTTTTTCCTGCAGTCCCTG, Exon 5-*ii* CTCCAAATGACTCCTGCCAT, Exon 6-*i* GCCTCTGCCATAGATTTTGC, Exon 6-*ii* TTCCTTCCCACCCTTTCTCT). Probes were created by random-primed labelling of PCR products with radioactive dCTP; these were used to screen a LANL chromosome 14-specific cosmid library (LA14NC01 obtained from the UK HGMP Resource Centre, Hinxton, Cambridge) using hybriding conditions set out in Church and Gilbert (1986). Positive clones (exon; LA1431-G5, exon 5: LA14123 - C4 and exon 6; LA1487 - D9) were prepared by a standard alkaline lysis protocol and taken through FISH analysis as above.

Results

Metaphase spreads from EBV-transformed cell lines were analysed by Fluorescence in situ Hybridisation (FISH) using successively smaller DNA probes (Fig.1). A breakpoint spanning BAC clone was obtained by FISH screening (RPCI-11 BAC 1078i14, acc. no. AL161851). EST sequences were examined in the genomic DNA flanking the breakpoint in order to identify potential transcripts in the locality. A number of ESTs were identified which had been annotated as containing homologous sequence to the conserved "PAS" domain present in a large number of genes (Gu et al, 2000). A search of such genes revealed that the most closely related gene encoded a mouse brain-expressed transcript, neuronal pas domain protein 3 (NPAS3 (MOP6), acc. no. AF137871; hereafter referred to as mNPAS3). Nucleotide homology to the mNPAS3 cDNA within human genomic DNA BAC clone sequences at 14q13 using the BLAST algorithm identified 12 exons corresponding to the human orthologue of mNPAS3 (*hNPAS3*) distributed over a genomic region of approximately 800-900Kb making it among the largest gene loci in the human genome (Figure 6). Subsequently, full length h*NPAS3* cDNA sequences have been submitted by two other groups to GenBank/EMBL with the accession numbers, AB054575 and AF164438, although these have differences to the mouse splice-form in the 5' exons. This is due to the presence of two alternative transcription start sites employed in both human and mouse genes. This was confirmed by analysis of published cDNA and EST sequences coupled

with further sequencing of corresponding IMAGE clones. These splice variants are highlighted in Figures 1, 3 and 6.

The ratio of fluorescent signals on the derived chromosomes 9 and 14 from the breakpoint-spanning BAC probe, 1078i14, indicated that the breakpoint was located at the centromeric end of the BAC. This is the location of exon 5 of the gene. Exon 4-, 5- and 6-containing cosmids were isolated and used as FISH probes to provide definitive proof of the location of the breakpoint and confirmation that a full-length transcript (and hence protein) cannot be synthesized on the derived chromosome 14. An exon 5-containing cosmid (see Figure 6) spanned the breakpoint. Subsequently a long-range PCR product-derived FISH probe corresponding to exon 5 indicated that the breakpoint lay upstream of exon 5.

Long-range PCR primers - NPAS3 exon 5

a) ccagcttgtatgtggtgtgg

b) ttactcccagtgcccattgt.

Discussion

A FISH-based approach has shown that the gene, *NPAS3*, is disrupted by a chromosomal rearrangement present in a mother and daughter who suffer from comorbid schizophrenia and learning disability respectively. NPAS3 is a brain expressed transcription factor of the basic helix-loop-helix PAS domain class which includes members such as AHR

and ARNT.

Neuronal pas3 (NPAS3) was originally cloned in the mouse (Brunskill et al, 1999) on the basis of its sequence homology with other PAS domain proteins. Its expression has been characterised in the developing mouse embryo where high levels are seen in the neural tube, neuroepithelium and, later, the neopallial layer of the cortex. Non-neural expression was also observed in the heart, limb and kidney. In the mouse, NPAS1 (human chromosomal location, 19q13) is expressed in deep pyramidal cortex cells, hippocampus and amygdala (Zhou et al., 1997). NPAS2 (human chromosomal location, 2q13) is expressed in the cortex, hippocampus and thalamus. Lower levels were also seen in spinal cord, intestines and uterus. NPAS2 was also recently deleted in mice by homologous recombination (Garcia et al., 2000) leading to deficits in cued and contextual memory. In addition NPAS2 appears to have a role in cellular energy state monitoring and the circadian rhythm pathway (Reick et al, 2001 and Rutter et al, 2001). The translocation event described herein disrupts the gene between exons 4 and 5. If transcription occurred at this disrupted locus, a truncated protein would result containing only the bHLH domain. It is conceivable that this protein would have a dominant negative effect on wild-type NPAS3 protein (or any other heterodimeric protein partner) through the creation of non-functional dimers (see Figure 7 for explanatory diagram). This would result in a potentially more severe or penetrant phenotype than a conventional point mutation.

Two examples where bHLH-PAS proteins have been altered through loss of the C-terminal PAS domain (one experimentally, the other in a patient with a chromosome translocation) have resulted in probable dominant negative action (Maemura et al, 1999, Holder jr. et al, 2000).

Mutations in this gene in karyotypically normal individuals would not be expected to have as severe or penetrant effects as those observed in the two t(9;14) patients.

Sequence comparison between hNPAS3 and other members of the NPAS sub-family show that homologies are largely restricted to the N-terminal end of the protein; the location of a basic helix-loop-helix and PAS domains. The greatest homology is with NPAS1, then NPAS2 and other PAS domain-containing proteins (data not shown). An alignment of the cognate human (conceptually translated from the splice-form containing exons 1-12) and mouse NPAS3 proteins reveals near identity over the N-terminal half of the protein but increased divergence at the C-terminal end. This is particularly the case for two stretches where 5 and 7 amino acids, respectively, have been gained in the human orthologue (Fig.4). These correspond to two poly-glycine tracts present within exon 12 (of 11 and 10 residues respectively). Such tracts can be indicative of slipped strand mispairing whereby trinucleotide repeats are aberrantly expanded or deleted. Where they occur in coding sequence, increases in the number of trinucleotide repeats can have a pathological effect on protein function (e.g.

Huntington disease and Spino-cerebellar ataxia 1). Another feature of such repeats is their unstable nature between generations: a lowering of the age of onset of a disease from generation to generation (anticipation) can often be directly linked to an increase in the number of repeat units.

Exon 12 (coding for the C-terminus of the protein) is also noteworthy because of the extremely high density of CpG dinucleotides (in humans and mouse); a feature that abruptly ends at the junctions with flanking intronic/3' sequences. This "CpG island" is unusual because it is both transcribed and also located at the 3' rather than 5' end of the gene. The significance of this in terms of potential transcriptional control by methylation or susceptibility to mutation is as yet unknown. However, the high level of G and C bases creates a bias in amino acid composition such that alanine, glycine, histidine and proline are over-represented. This may explain the presence and expansion of the poly-glycine tracts in Npas3.

14q13 is also the site of linkage to Fahr's syndrome (idiopathic basal ganglia calcification; IBGC) as determined from analysis of families (Geschwind et al, 1999). Fahr's syndrome symptoms are often accompanied by psychoses such as schizophrenia. Thus, it may be the case that NPAS3 is also the gene responsible for Fahr's syndrome.

REFERENCES

1. Berrettini W.H. (2000) Genetics of psychiatric disease. Annu. Rev. Med. 51:465 - 79.

2. Lai C.S, Fisher S.E, Hurst J.A, Vargha-Khadem F, Monaco A.P. (2001) A forkhead-domain gene is mutated in a severe speech and language disorder. Nature. 413(6855):519 - 23.

3. Kleinjan D.J, van Heyningen V. (1998) Position effect in human genetic disease. Hum Mol Genet. 7(10):1611 - 8.

4. Brennan J, Skarnes W.C. (1999) Gene trapping in mouse embryonic stem cells. Methods Mol. Biol. 97:123 - 38.

5. Breen M, Arveiler B, Murray I, Gosden J.R, Porteous D.J. (1992) YAC mapping by FISH using Alu-PCR-generated probes. Genomics. 13(3):726 - 30.

6. Gu Y.Z, Hogenesch J.B, Bradfield C.A. The PAS superfamily: sensors of environmental and developmental signals. (2000) Annu Rev Pharmacol Toxicol.;40:519 - 61.

7. Brunskill E.W, Witte D.P, Shreiner A.B, Potter S.S. (1999) Characterization of npas3, a novel basic helix-loop-helix PAS gene expressed in the developing mouse nervous system. Mech Dev. Nov;88(2):237 - 41.

8. Zhou Y.D, Barnard M, Tian H, Li X, Ring H.Z, Francke U, Shelton J, Richardson J, Russell D.W, McKnight S.L. (1997) Molecular characterization of two mammalian bHLH-PAS domain proteins selectively expressed in the central nervous system. Proc Natl Acad Sci U S A. Jan 21;94(2):713 -8.

9. Garcia J.A, Zhang D, Estill S.J, Michnoff C, Rutter J, Reick M, Scott K, Diaz-Arrastia R, McKnight S.L. (2000) Impaired cued and contextual memory in NPAS2-deficient mice. Science. Jun 23;288(5474):2226 - 30.

10. Reick M, Garcia J.A, Dudley C, McKnight S.L. (2001) NPAS2: an analog of clock operative in the mammalian forebrain. Science. Jul 20;293(5529):506 - 9.

11. Rutter J, Reick M, Wu L.C, McKnight S.L. (2001) Regulation of clock and NPAS2 DNA binding by the redox state of NAD cofactors. Science. Jul 20;293(5529):510 - 4.

12. Maemura K, Hsieh C.M, Jain M.K, Fukumoto S, Layne M.D, Liu Y, Kourembanas S, Yet S.F, Perrella M.A, Lee M.E. (1999) Generation of a dominant-negative mutant of endothelial PAS domain protein 1 by deletion of a potent C-terminal transactivation domain. J Biol Chem. Oct 29;274(44):31565 - 70.

13. Holder J.L Jr, Butte N.F, Zinn A.R. (2000) Profound obesity associated with a balanced translocation that disrupts the SIM1 gene. Hum Mol Genet. Jan 1;9(1):101 - 8.

14. Geschwind D.H, Loginov M, Stern J.M. (1999) Identification of a locus on chromosome 14q for idiopathic basal ganglia calcification (Fahr disease). Am J Hum Genet. 1999 Sep;65(3):764 - 72.

Figure 1

NPAS3 (NM_022123) nucleic acid sequence (spliceform 1b-3-4etc)

```
   1 ccacgcgtcc gacgcccccc acccgggagg ggggagagag gcaaaaagta agagaggaaa
  61 aaaaatagca ggaagatggc gcccaccaag cccagctttc agcaggatcc ttccaggcga
 121 gaacgtttac aagcattgag aaaggagaaa tcccgagatg ctgctcgctc ccgccgggga
 181 aaagaaaact ttgagttcta tgaattggcc aagttgttgc ctcttcctgc agccattacc
 241 agccagctcg acaaggcatc catcattcga cttacaatta gctatctgaa aatgagggac
 301 tttgctaacc agggggaccc tccgtggaac ttgcgaatgg aaggccctcc acctaacaca
 361 tcagtaaaag gtgcacagcg aaggagaagc cccagtgcac tagccattga agtatttgaa
 421 gcacatttgg gaagccacat tttgcagtcc ctggatggct ttgtatttgc actaaatcag
 481 gaaggaaaat ttttgtacat ttccgaaaca gtctccatct acctaggcct ctcacaagtg
 541 gagctgacag gcagcagtgt ctttgactat gtccaccccg gagatcacgt ggagatggct
 601 gagcagctgg gcatgaagct ccccccctggg cggggtctcc tgtcacaggg cactgctgag
 661 gacggagcca gctcagcatc ttcctcctct cagtcggaga cccccgagcc agtggagtca
 721 accagcccca gtctgctaac cactgacaac actcttgagc gttccttttt catccgaatg
 781 aaatctactc tgaccaaacg cggtgtgcac atcaaatcat caggatataa ggtgattcac
 841 ataacaggcc ggctacgcct gagagtgtcg ctgtcccacg ggaggaccgt ccccagccaa
 901 atcatgggtc tcgtggttgt tgcgcatgcc ttgcctcccc ctacgatcaa tgaagtcaga
 961 attgactgcc atatgttcgt cactcgagta aatatggacc tcaatatcat ttactgtgaa
1021 aataggatta gtgattatat ggatctgacc cctgtagata tcgtagggaa gagatgctac
1081 cacttcatcc atgctgaaga cgtggagggc atcaggcaca gtcacttgga cttgctgaat
1141 aagggtcagt gtgtgacaaa gtactatcgc tggatgcaga agaacggagg atatatttgg
1201 atacagtcca gtgccaccat agctattaat gccaagaatg caaatgaaaa gaatatcatc
1261 tgggtgaatt accttcttag caatcctgag tacaaggaca cacccatgga catcgcacag
1321 ctcccccatc tgccggagaa aacttccgaa tcctcggaga catccgactc tgagtcagac
1381 tctaaagaca cctcaggtat tacagaggac aacgagaact ccaagtccga cgagaagggg
1441 aaccagtccg agaacagcga agacccggag cccgaccgga agaagtcggg caacgcgtgt
1501 gacaacgaca tgaactgcaa cgacgacggc cacagctcca gtaacccgga cagccgcgac
1561 agcgacgaca gcttcgagca ctcggacttt gagaacccca aggcgggcga ggacggcttc
1621 ggtgctctgg gcgcgatgca gatcaaggtg gagcgctacg tggagagcga gtcggacctg
1681 cggctgcaga actgcgagtc actcacgtcc gacagcgcca aggactcgga cagcgcaggc
1741 gaggcgggcg cgcaggcctc cagcaagcac cagaagcgca agaaaaggcg gaaacggcaa
1801 aagggcggca gcgccagccg ccggcgcctg tccagcgcgt cgagcccagg cggcctggac
1861 gcgggcctgg tggagccccc gcggctgctg tcctccccca acagtgcctc ggtgctcaag
1921 atcaagacgg agatctcaga acccatcaat ttcgacaatg acagcagcat ctggaactac
1981 ccgcccaacc gggagatctc caggaacgag tccccctaca gcatgaccaa gccccccagc
2041 tctgagcact tcccgtcccc gcagggcggc ggcggtgggg gtggcggtgg cgggggctg
2101 cacgtggcca ttcccgactc ggtcctcacc ccgcccggcg ccgacggcgc ggccgcccgc
2161 aagactcagt tcggcgcctc ggccaccgcg gccctggccc ccgtcgcctc cgacccgctg
2221 tcacccccgc tctcggcgtc cccgcgggac aagcaccccg ggaacggcgg cgggggcggg
2281 ggcgggggcg gcggcgcggg gggcggcggc cccagcgcgt ccaactcctt gctgtacact
2341 ggggacctgg aggcgctgca gaggttgcag gcgggcaacg tcgtgctccc gctggtgcac
2401 agggtgaccg ggaccctggc cgccaccagc acggccgcgc agagggtcta caccacgggc
2461 accatccgct acgcgcccgc cgaggtgacc ctggccatgc agagcaacct gctgcccaac
2521 gcgcacgctg ttaacttcgt ggacgttaac agccccggct ttggcctcga ccccaagacg
2581 cccatggaga tgctctacca ccacgtgcac cggctcaaca tgtcaggacc gttcggcggc
2641 gcagtgagcg cagctagcct gacgcagatg cccgccggca acgtgttcac cacggccgag
2701 ggactcttct ccacgctgcc cttccccgtc tacagcaacg gcatccacgc ggcacagact
2761 ctggagcgca aggaggactg aggcgccgcc cgtcctgggc ccggccaggc cccgcttgga
2821 ggaggcatcg tcggcatttt cgtttagacc tttaattcta gcactttgaa ttcgagcagg
2881 tcagcgtctt ctctcgccac gacggtcccc attccacccc ctctttcttt cacctgactt
2941 attctttcgt gtaaagatat gtttattttt tgccttcaga gggtcagacg accagttgcc
3001 tgccgttttg tcttcttcta aggtgtgtgt tgggttgttt tgctttcctt tgcatcttta
3061 ttaagatgtc tttcatgtgt atatgcctct gccatagaat actcagtctt gtggtcaaga
3121 gagttctcaa gtgacaacca ttggggtttc ttcataaaga tcttgatatg atcaagatgg
3181 aaagagacaa gcataaacaa tgtgccctgt ttgactaagt caaatgaaat agggtggttt
3241 ttgtttctgt tcctaattcc tttaaaaaat agggggaata gtattttaga attttatgca
3301 gaatttaatt ctcttttac ggttaagatt ttaagatttt cttacttgca cataaaaata
3361 atttgggttc ttaaacttaa tttctggcct gtgactagaa tgtttaaaaa aaaaaaaaac
3421 cctcgtgc
```

Figure 2

NPAS3 protein sequence (spliceform 1b-3-4etc.)

[redacted]QALRKEKSRDAARSRRGKENFEFYELAKLLPLPAAITSQ
LDKASIIRLTISYLKMRDFANQGDPPWNLRMEGPPPNTSVKGAQRRRSPSALAIEVF
EAHLGSHILQSLDGFVFALNQEGKFLYISETVSIYLGLSQVELTGSSVFDYVHPGDH
VEMAEQLGMKLPPGRGLLSQGTAEDGASSASSSSQSETPEPVESTSPSLLTTDNTLE
RSFFIRMKSTLTKRGVHIKSSGYKVIHITGRLRLRVSLSHGRTVPSQIMGLVVVAHA
LPPPTINEVRIDCHMFVTRVNMDLNIIYCENRISDYMDLTPVDIVGKRCYHFIHAED
VEGIRHSHLDLLNKGQCVTKYYRWMQKNGGYIWIQSSATIAINAKNANEKNIIWVNY
LLSNPEYKDTPMDIAQLPHLPEKTSESSETSDSESDSKDTSGITEDNENSKSDEKGN
QSENSEDPEPDRKKSGNACDNDMNCNDDGHSSSNPDSRDSDDSFEHSDFENPKAGED
GFGALGAMQIKVERYVESESDLRLQNCESLTSDSAKDSDSAGEAGAQASSKHQKRKK
RRKRQKGGSASRRRLSSASSPGGLDAGLVEPPRLLSSPNSASVLKIKTEISEPINFD
NDSSIWNYPPNREISRNESPYSMTKPPSSEHFPSPQGGGGGGGGGGGLHVAIPDSVL
TPPGADGAAARKTQFGASATAALAPVASDPLSPPLSASPRDKHPGNGGGGGGGGGGA
GGGGPSASNSLLYTGDLEALQRLQAGNVVLPLVHRVTGTLAATSTAAQRVYTTGTIR
YAPAEVTLAMQSNLLPNAHAVNFVDVNSPGFGLDPKTPMEMLYHHVHRLNMSGPFGG
AVSAASLTQMPAGNVFTTAEGLFSTLPFPVYSNGIHAAQTLERKED

Figure 3

NPAS3 nucleic acid sequence (spliceform incorporating exons 1a-2-3-4etc) similar to mouse cDNA with accession number NM_013780)

```
   1 ATGGGGAGGG CCGGCGCCGC GGCCAACGGC ACCCCGCAGA ACGTCCAGGG CATCACCTCC
  61 TACCAGCAGC GAATAACTGC CCAGCATCCT CTGCCCAACC AATCAGAATG TAGGAAAATC
 121 TACAGATATG ACGGAATCTA CTGTGAATCT ACCTACCAGA ATTTACAAGC ATTGAGAAAG
 181 GAGAAATCCC GAGATGCTGC TCGCTCCCGC CGGGGAAAAG AAAACTTTGA GTTCTATGAA
 241 TTGGCCAAGT TGTTGCCTCT TCCTGCAGCC ATTACCAGCC AGCTCGACAA GGCATCCATC
 301 ATTCGACTTA CAATTAGCTA TCTGAAAATG AGGGACTTTG CTAACCAGGG GGACCCTCCG
 361 TGGAACTTGC GAATGGAAGG CCCTCCACCT AACACATCAG TAAAAGGTGC ACAGCGAAGG
 421 AGAAGCCCCA GTGCACTAGC CATTGAAGTA TTTGAAGCAC ATTTGGGAAG CCACATTTTG
 481 CAGTCCCTGG ATGGCTTTGT ATTTGCACTA AATCAGGAAG GAAAATTTTT GTACATTTCC
 541 GAAACAGTCT CCATCTACCT AGGCCTCTCA CAAGTGGAGC TGACAGGCAG CAGTGTCTTT
 601 GACTATGTCC ACCCCGGAGA TCACGTGGAG ATGGCTGAGC AGCTGGGCAT GAAGCTCCCC
 661 CCTGGGCGGG GTCTCCTGTC ACAGGGCACT GCTGAGGACG GAGCCAGCTC AGCATCTTCC
 721 TCCTCTCAGT CGGAGACCCC CGAGCCAGTG GAGTCAACCA GCCCCAGTCT GCTAACCACT
 781 GACAACACTC TTGAGCGTTC CTTTTTCATC CGAATGAAAT CTACTCTGAC CAAACGCGGT
 841 GTGCACATCA AATCATCAGG ATATAAGGTG ATTCACATAA CAGGCCGGCT ACGCCTGAGA
 901 GTGTCGCTGT CCCACGGGAG GACCGTCCCC AGCCAAATCA TGGGTCTCGT GGTTGTTGCG
 961 CATGCCTTGC CTCCCCCTAC GATCAATGAA GTCAGAATTG ACTGCCATAT GTTCGTCACT
1021 CGAGTAAATA TGGACCTCAA TATCATTTAC TGTGAAAATA GGATTAGTGA TTATATGGAT
1081 CTGACCCCTG TAGATATCGT AGGGAAGAGA TGCTACCACT TCATCCATGC TGAAGACGTG
1141 GAGGGCATCA GGCACAGTCA CTTGGACTTG CTGAATAAGG GTCAGTGTGT GACAAAGTAC
1201 TATCGCTGGA TGCAGAAGAA CGGAGGATAT ATTTGGATAC AGTCCAGTGC CACCATAGCT
1261 ATTAATGCCA AGAATGCAAA TGAAAAGAAT ATCATCTGGG TGAATTACCT TCTTAGCAAT
1321 CCTGAGTACA AGGACACACC CATGGACATC GCACAGCTCC CCCATCTGCC GGAGAAAACT
1381 TCCGAATCCT CGGAGACATC CGACTCTGAG TCAGACTCTA AAGACACCTC AGGTATTACA
1441 GAGGACAACG AGAACTCCAA GTCCGACGAG AAGGGGAACC AGTCCGAGAA CAGCGAAGAC
1501 CCGGAGCCCG ACCGGAAGAA GTCGGGCAAC GCGTGTGACA ACGACATGAA CTGCAACGAC
1561 GACGGCCACA GCTCCAGTAA CCCGGACAGC CGCGACAGCG ACGACAGCTT CGAGCACTCG
1621 GACTTTGAGA ACCCCAAGGC GGGCGAGGAC GGCTTCGGTG CTCTGGGCGC GATGCAGATC
1681 AAGGTGGAGC GCTACGTGGA GAGCGAGTCG GACCTGCGGC TGCAGAACTG CGAGTCACTC
1741 ACGTCCGACA GCGCCAAGGA CTCGGACAGC GCAGGCGAGG CGGGCGCGCA GGCCTCCAGC
1801 AAGCACCAGA AGCGCAAGAA AAGGCGGAAA CGGCAAAAGG GCGGCAGCGC CAGCCGCCGG
1861 CGCCTGTCCA GCGCGTCGAG CCCAGGCGGC CTGGACGCGG GCCTGGTGGA GCCCCCGCGG
1921 CTGCTGTCCT CCCCCAACAG TGCCTCGGTG CTCAAGATCA AGACGGAGAT CTCAGAACCC
1981 ATCAATTTCG ACAATGACAG CAGCATCTGG AACTACCCGC CCAACCGGGA GATCTCCAGG
2041 AACGAGTCCC CCTACAGCAT GACCAAGCCC CCCAGCTCTG AGCACTTCCC GTCCCCGCAG
2101 GGCGGCGGCG GTGGGGGTGG CGGTGGCGGG GGGCTGCACG TGGCCATTCC CGACTCGGTC
2161 CTCACCCCGC CCGGCGCCGA CGGCGCGGCC GCCCGCAAGA CTCAGTTCGG CGCCTCGGCC
2221 ACCGCGGCCC TGGCCCCCGT CGCCTCCGAC CCGCTGTCAC CCCCGCTCTC GGCGTCCCCG
2281 CGGGACAAGC ACCCCGGGAA CGGCGGCGGG GGCGGGGGCG GGGGCGGCGG CGCGGGGGGC
2341 GGCGGCCCCA GCGCGTCCAA CTCCTTGCTG TACACTGGGG ACCTGGAGGC GCTGCAGAGG
2401 TTGCAGGCGG GCAACGTCGT GCTCCCGCTG GTGCACAGGG TGACCGGGAC CCTGGCCGCC
2461 ACCAGCACGG CCGCGCAGAG GGTCTACACC ACGGGCACCA TCCGCTACGC GCCCGCCGAG
2521 GTGACCCTGG CCATGCAGAG CAACCTGCTG CCCAACGCGC ACGCTGTTAA CTTCGTGGAC
2581 GTTAACAGCC CCGGCTTTGG CCTCGACCCC AAGACGCCCA TGGAGATGCT CTACCACCAC
2641 GTGCACCGGC TCAACATGTC AGGACCGTTC GGCGGCGCAG TGAGCGCAGC TAGCCTGACG
2701 CAGATGCCCG CCGGCAACGT GTTCACCACG GCCGAGGGAC TCTTCTCCAC GCTGCCCTTC
2761 CCCGTCTACA GCAACGGCAT CCACGCGGCA CAGACTCTGG AGCGCAAGGA GGACTGAGGC
2821 GCCGCCCGTC CTGGGCCCGG CCAGGCCCCG CTTGGAGGAG GCATCGTCGG CATTTTCGTT
2881 TAGACCTTTA ATTCTAGCAC TTTGAATTCG AGCAGGTCAG CGTCTTCTCT CGCCACGACG
2941 GTCCCCATTC CACCCCCTCT T
```

Figure 4

NPAS3 protein sequence of spliceform incorporating exons 1a-2-3-4etc.

[redacted]QA
LRKEKSRDAARSRRGKENFEFYELAKLLPLPAAITSQLDKASIIRLTISYLKMRDFA
NQGDPPWNLRMEGPPPNTSVKGAQRRRSPSALAIEVFEAHLGSHILQSLDGFVFALN
QEGKFLYISETVSIYLGLSQVELTGSSVFDYVHPGDHVEMAEQLGMKLPPGRGLLSQ
GTAEDGASSASSSSQSETPEPVESTSPSLLTTDNTLERSFFIRMKSTLTKRGVHIKS
SGYKVIHITGRLRLRVSLSHGRTVPSQIMGLVVVAHALPPPTINEVRIDCHMFVTRV
NMDLNIIYCENRISDYMDLTPVDIVGKRCYHFIHAEDVEGIRHSHLDLLNKGQCVTK
YYRWMQKNGGYIWIQSSATIAINAKNANEKNIIWVNYLLSNPEYKDTPMDIAQLPHL
PEKTSESSETSDSESDSKDTSGITEDNENSKSDEKGNQSENSEDPEPDRKKSGNACD
NDMNCNDDGHSSSNPDSRDSDDSFEHSDFENPKAGEDGFGALGAMQIKVERYVESES
DLRLQNCESLTSDSAKDSDSAGEAGAQASSKHQKRKKRRKRQKGGSASRRRLSSASS
PGGLDAGLVEPPRLLSSPNSASVLKIKTEISEPINFDNDSSIWNYPPNREISRNESP
YSMTKPPSSEHFPSPQGGGGGGGGGGGLHVAIPDSVLTPPGADGAAARKTQFGASAT
AALAPVASDPLSPPLSASPRDKHPGNGGGGGGGGGGAGGGGPSASNSLLYTGDLEAL
QRLQAGNVVLPLVHRVTGTLAATSTAAQRVYTTGTIRYAPAEVTLAMQSNLLPNAHA
VNFVDVNSPGFGLDPKTPMEMLYHHVHRLNMSGPFGGAVSAASLTQMPAGNVFTTAE
GLFSTLPFPVYSNGIHAAQTLERKED

Figure 5 Representation of the balanced translocation disrupting the NPAS3 gene on chromosome 14

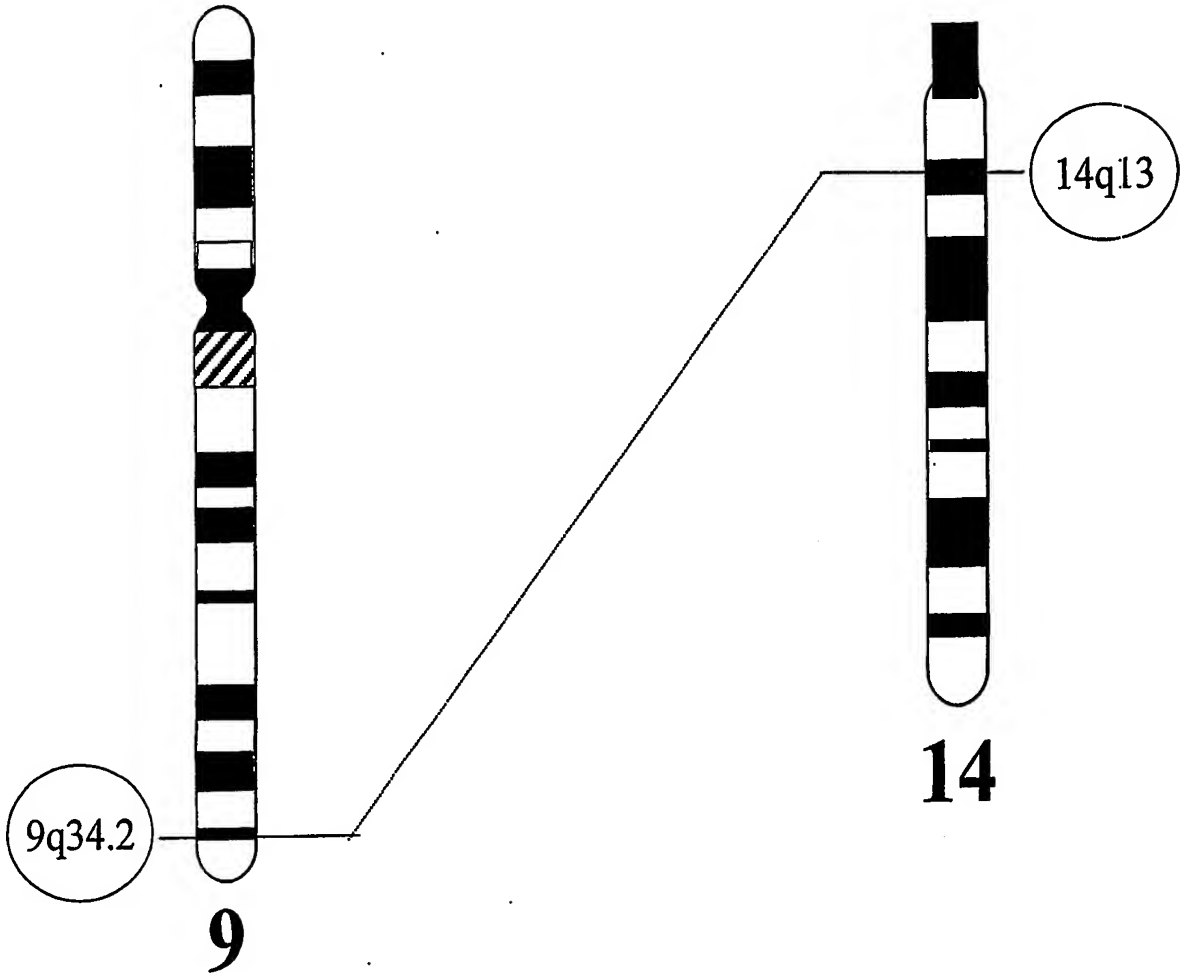
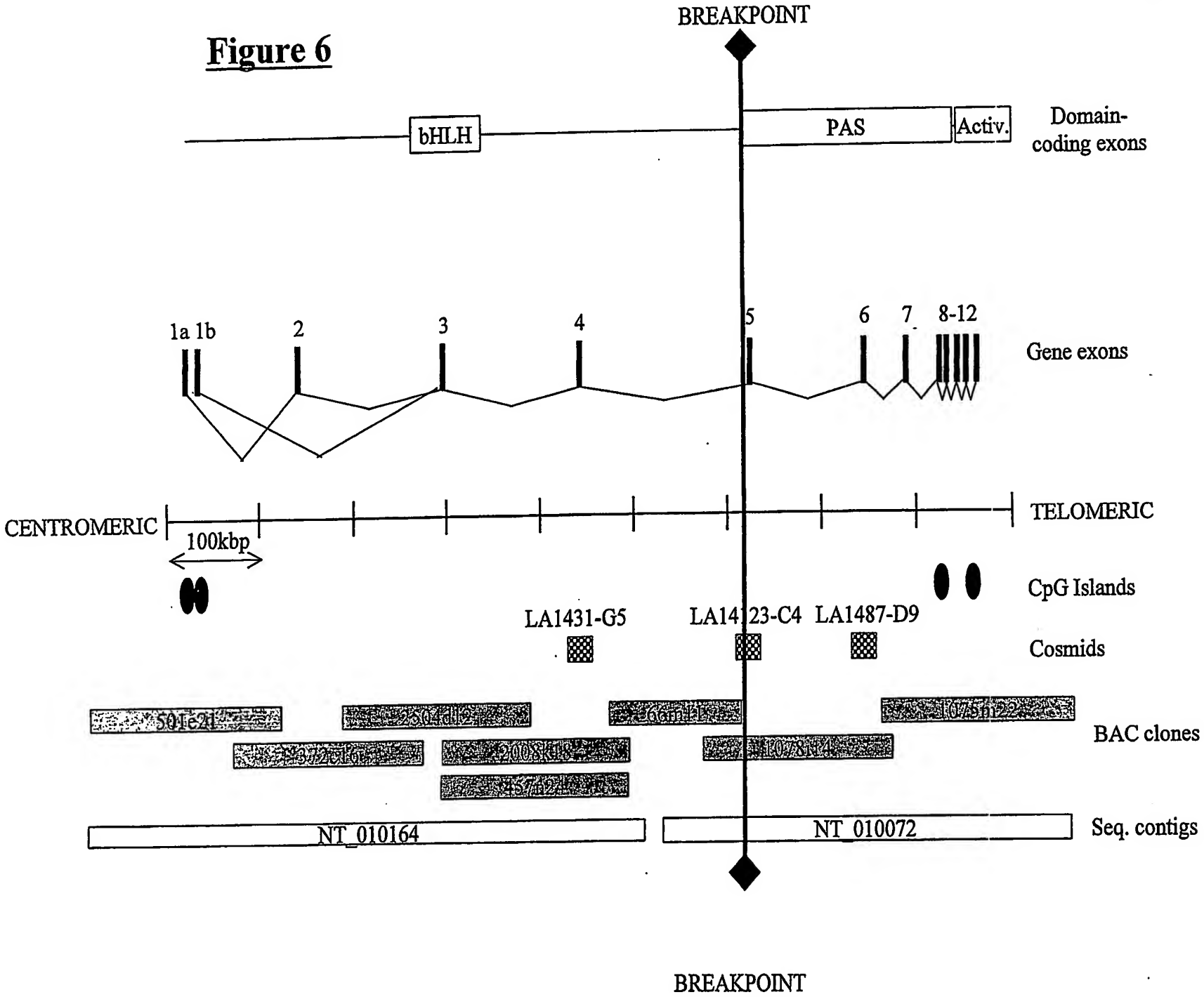


Figure 6


BREAKPOINT
bHLH
PAS
Activ.
Domain-coding exons
1a 1b
2
3
4
5
6
7
8-12
Gene exons
CENTROMERIC
TELOMERIC
100kbp
CpG Islands
LA1431-G5
LA1423-C4
LA1487-D9
Cosmids
BAC clones
NT_010164
NT_010072
Seq. contigs
BREAKPOINT

The position of the breakpoint may cause the production of a truncated polypeptide with dominant negative activity

Figure 7